

Understanding Society User Support - Support #700

longitudinal weights for small sub-samples

01/13/2017 01:01 PM - David Bartram

Status:	Closed	Start date:	01/13/2017
Priority:	High	% Done:	100%
Assignee:	David Bartram		
Category:	Weights		
Description			
<p>Some questions re longitudinal weights. Having read "Weighting Strategy for Understanding Society", I gather that absence from a single wave leads to longitudinal weights of 0 for all later waves. In my case, this means that I am losing more than 10% of the small (sub)sample (n=997) I want to analyze. For context, my analysis starts with people who were non-citizens in Wave 1 and then revisits them in Wave 6 (by which point some have become UK citizens).</p> <p>To avoid losing so many cases I'm wondering about replacing the zero values with some reasonable substitute. Three possibilities, perhaps: a) the mean of <code>f_indinus_lw</code> for the sub-sample (as calculated -- here, 0.7254; b) the cross-sectional individual weight at Wave 6, <code>f_indinui_xw</code>; or c), a cross-sectional individual weight at Wave 1. No doubt each is sub-optimal, but losing >10% of the sample is also sub-optimal. Any comments as to the relative merits of these three options? (Or are they all bad...)</p> <p>If I could assume that attrition is not related to the response variable, then perhaps instead use the cross-sectional weight from Wave 1 for everyone? Response variables are life satisfaction (<code>scfssato</code>), interest in politics (<code>Vote6</code>), and importance of British identity (<code>britid</code>). Obviously it's up to me to make some sort of informed choice about this, but I'd be grateful for any comment.</p> <p>One additional point, perhaps important for context. Because <code>svy</code> doesn't work with <code>xt-</code> commands, I can't use <code>subpop</code> -- so, I have assigned a weight of 0 to all those not in the subpopulation of interest (i.e., all but the 997). Is that the correct approach for using e.g. <code>xtlogit</code>? The mean of the weights for the small subpopulation is then no longer 1; do the weights nonetheless ensure that the subsample is (reasonably) representative of the subpopulation?</p>			

History

#1 - 01/13/2017 01:35 PM - David Bartram

Correction -- the mean of `f_indinus_lw` for the sub-sample is 1.0347, not 0.7254. (I had been including the zero values in the calculation of the mean -- ugh...) Using only those in the sub-sample who have non-zero values, the mean is 1.0347.

#2 - 01/16/2017 09:00 AM - Victoria Nolan

- *Category set to Weights*
- *Status changed from New to In Progress*
- *Assignee set to David Bartram*
- *% Done changed from 0 to 10*
- *Private changed from Yes to No*

Dear David,

Many thanks for your query. I have passed this on to our weighting team who will look into it for you.

Best wishes, Victoria.

On behalf of the Understanding Society User Support Team

#3 - 01/19/2017 02:35 PM - Peter Lynn

Hello David.

To my mind the best solution would be for you to use `a_indinus_xw` as a base weight and then multiply it by an adjustment factor for loss from the sample by `w6`. You would have to calculate this factor by fitting a model (e.g. logit) based on all wave 1 respondents in your subgroup of interest, in which the dep var is a 0/1 indicator of whether they also responded at `w6` (and removing from the base any known to have died or emigrated before `w6`). Predictor variables can be anything relevant observed at `w1`. This will give you a predicted probability for every `w1` respondent of responding at `w6`. Call this `P`. You then need to adjust `a_indinus_xw` by multiplying it by `1/P` for all the cases that can be included in your analysis.

An amended version of your a) might be second-best option. Instead of just taking the mean, take the mean within groups defined by relevant (to your analysis) variables. It sounds like you have 800+ cases with a non-zero weight (and 100 or so with zero?) so you have a big enough sample to divide into a good number of groups (10 to 20?)

Do not use approach b), as that would be very distorting due to the inclusion of the new boost sample in the "ui" weights, but not in your analysis (i.e. all ethnic minorities and immigrants will be greatly weighted down - much more than they should be).

And I wouldn't recommend c) either, as I would doubt that 5 years of attrition is ignorable.

Peter

#4 - 01/19/2017 02:44 PM - Victoria Nolan

- *Status changed from In Progress to Feedback*

- *% Done changed from 10 to 80*

#5 - 01/23/2017 12:45 PM - David Bartram

Thank you Peter -- that's a very helpful response.

#6 - 01/30/2017 11:51 AM - Victoria Nolan

- *Status changed from Feedback to Closed*

- *% Done changed from 80 to 100*